Comparative Effectiveness Review
Number 143

Effective Health Care Program

Imaging Techniques for the Diagnosis and Staging of Hepatocellular Carcinoma

Executive Summary

Background and Objectives

Hepatocellular carcinoma (HCC) is the most common primary malignant neoplasm of the liver, usually developing in individuals with chronic liver disease or cirrhosis. Worldwide, it is the fifth most common cancer and the third most common cause of cancer death.[1] There were 156,940 deaths attributed to liver and intrahepatic bile duct cancer in the United States in 2011, with 221,130 new cases diagnosed.[2] The lifetime risk of developing liver and intrahepatic bile duct cancer in the United States is about 1 in 132, with an age-adjusted incidence rate of 7.3 per 100,000 people per year.[3]

The American Association for the Study of Liver Diseases (AASLD) recommends surveillance for the following groups at high risk for developing HCC: Asian male hepatitis B virus (HBV) carriers age 40 and older, Asian female HBV carriers age 50 and older, HBV carriers with a family history of HCC, African/North American Black HBV carriers, HBV or hepatitis C virus carriers with cirrhosis, all individuals with other causes for cirrhosis (including alcoholic cirrhosis), and patients with stage 4 primary biliary cirrhosis.[4]

The natural history of HCC is variable, but it is often an aggressive tumor associated with poor survival without treatment.[5]

Effective Health Care Program

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at **www.effectivehealthcare.ahrq.gov/reports/final.cfm**.

When diagnosed early, HCC may be amenable to potentially curative therapy. The three phases of pretherapy evaluation of HCC are detection, further evaluation of focal liver lesions, and staging.[4] Detection

AHRQ
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Effective Health Care

often occurs in the setting of surveillance or in the use of periodic testing in people without HCC to identify lesions in the liver that are clinically suspicious for HCC.[4] The evaluation phase involves the use of additional tests (radiological and/or histopathological) to confirm that a focal liver lesion is indeed HCC. Staging determines the extent and severity of a person's cancer to inform prognosis and treatment decisions. A number of staging systems are available, including the widely used TNM (tumor, node, metastasis) staging system and the more recent Barcelona Clinic Liver Cancer (BCLC) staging system,[6] which has become the de facto staging reference standard;[4] the Milan criteria have been used to identify patients likely to experience better post-transplantation outcomes, although other methods have been proposed.[7]

A number of imaging techniques are available to detect the presence of lesions, evaluate focal liver lesions, and determine the stage of the disease. They include ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET). Understanding the diagnostic accuracy of imaging methods and how they affect clinical decisionmaking, and ultimately patient outcomes, is a challenge. Imaging techniques may be used alone, in various combinations or algorithms, and/or with liver-specific biomarkers, resulting in many potential comparisons. Technical aspects of imaging methods are complex, and they are continuously evolving.

Diagnostic accuracy studies use different reference standards, such as explanted liver specimens from patients undergoing transplantation, percutaneous or surgical biopsy, imaging, clinical followup, or combinations of these methods. Use of these different reference standards introduces heterogeneity that may limit comparisons of techniques. Reference standards also are susceptible to misclassification due to sampling error, inadequate specimens, insufficient followup, or other factors. Other considerations may impact the diagnostic accuracy or clinical utility of imaging strategies; they include risk factors for HCC and lesion characteristics, such as tumor size or degree of differentiation, severity of hepatic fibrosis, and etiology of liver disease.

Accurate diagnosis and staging of HCC are critical for providing optimal patient care. However, clinical uncertainty remains regarding optimal imaging strategies due to the factors described above. The purpose of this report is to comprehensively review the comparative effectiveness and diagnostic performance of different imaging modalities and strategies for detection of HCC, evaluation of focal liver lesions to identify HCC, and staging of HCC.

Scope and Key Questions

The Key Questions and corresponding analytic frameworks used to guide this report are shown below. Separate analytic frameworks address detection (Figure A), diagnosis (Figure B), and staging (Figure C). The analytic frameworks show the target populations, interventions (imaging tests), and outcomes (diagnostic accuracy, clinical decisionmaking, clinical outcomes, and harms) that we examined.

Key Question 1. What is the comparative effectiveness of available imaging-based strategies, used singly or in sequence, for detection of hepatocellular carcinoma among individuals in surveillance and nonsurveillance settings?

a. What is the comparative test performance of imaging-based strategies for detecting HCC?
   - i How is a particular technique's test performance modified by use of various reference standards (e.g., explanted liver samples, histological diagnosis, or clinical and imaging followup)?
   - ii. How is the comparative effectiveness modified by patient (e.g., severity of liver disease, underlying cause of liver disease, body mass index, age, sex, race), tumor (e.g., tumor diameter, degree of differentiation, location), technical, or other factors (e.g., results of biomarker tests, setting)?

b. What is the comparative effectiveness of imaging-based strategies on intermediate outcomes related to clinical decisionmaking (e.g., use of subsequent diagnostic tests and treatments)?

c. What is the comparative effectiveness of imaging-based strategies on clinical and patient-centered outcomes?

d. What are the adverse effects or harms associated with imaging-based surveillance strategies?

Key Question 2. What is the comparative effectiveness of imaging techniques, used singly, in combination, or in sequence, in diagnosing hepatocellular carcinoma among individuals in whom a focal liver lesion has been detected?

a. What is the comparative test performance of imaging techniques for diagnosing HCC in patients with a focal liver lesion?
   - i. How is a particular technique's test performance modified by use of various reference standards (e.g., explanted liver samples, histological diagnosis, or clinical imaging and followup)?
   - ii. How is the comparative effectiveness modified by patient, tumor, technical, or other factors?

b. What is the comparative effectiveness of the various imaging techniques on intermediate outcomes related to clinical decisionmaking?

c. What is the comparative effectiveness of the various imaging techniques on clinical and patient-centered outcomes?

d. What are the adverse effects or harms associated with imaging-based diagnostic strategies?

Key Question 3. What is the comparative effectiveness of imaging techniques, used singly, in combination, or in sequence, in staging hepatocellular carcinoma among patients diagnosed with hepatocellular carcinoma?

a. What is the comparative test performance of imaging techniques to predict HCC tumor stage?

i. How is a particular technique's test performance modified by use of various reference standards (e.g., explanted liver samples, histological diagnosis, or clinical and imaging followup)?

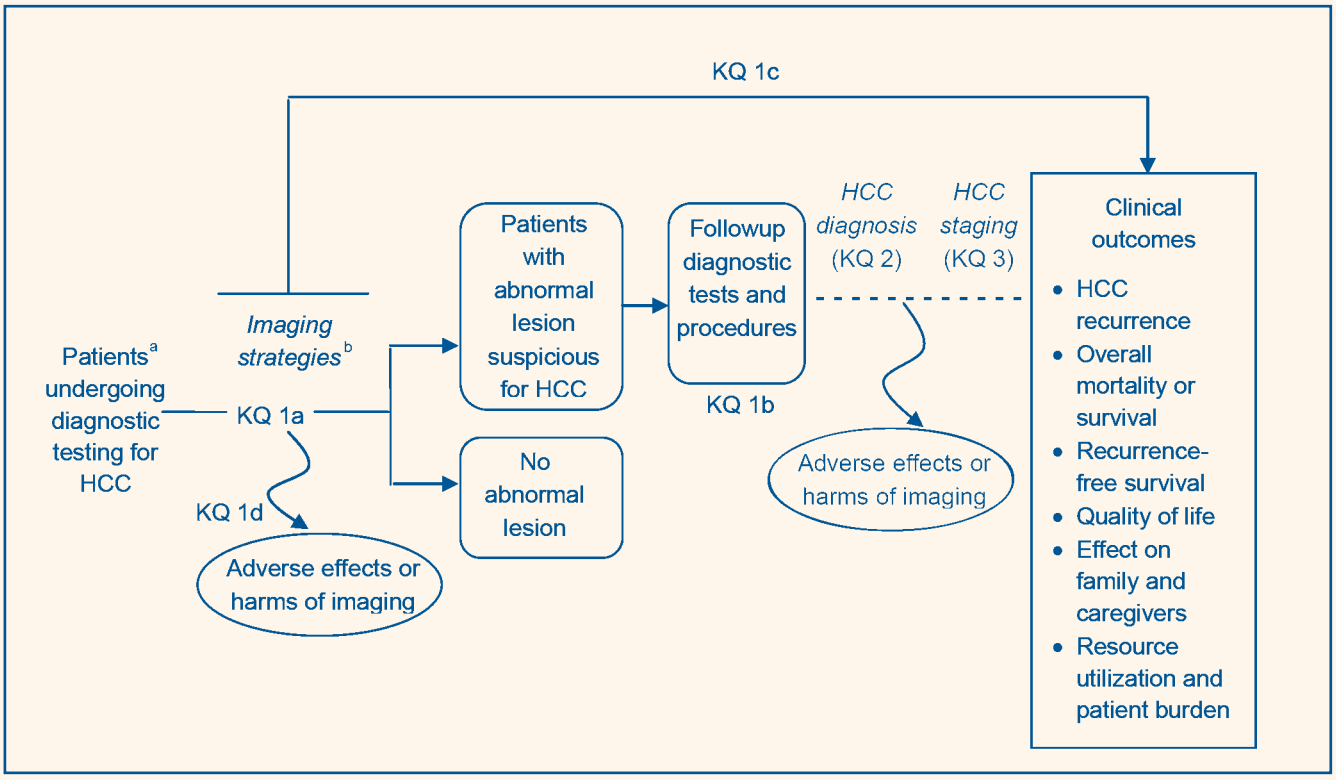
ii. How is the comparative effectiveness modified by patient, tumor, technical, or other factors?

b. What is the comparative test performance effectiveness of imaging techniques on intermediate outcomes related to clinical decisionmaking?

c. What is the comparative effectiveness of imaging techniques on clinical and patient-centered outcomes?

d. What are the adverse effects or harms associated with imaging-based staging strategies?

Figure A. Analytic framework—detection (Key Question 1)



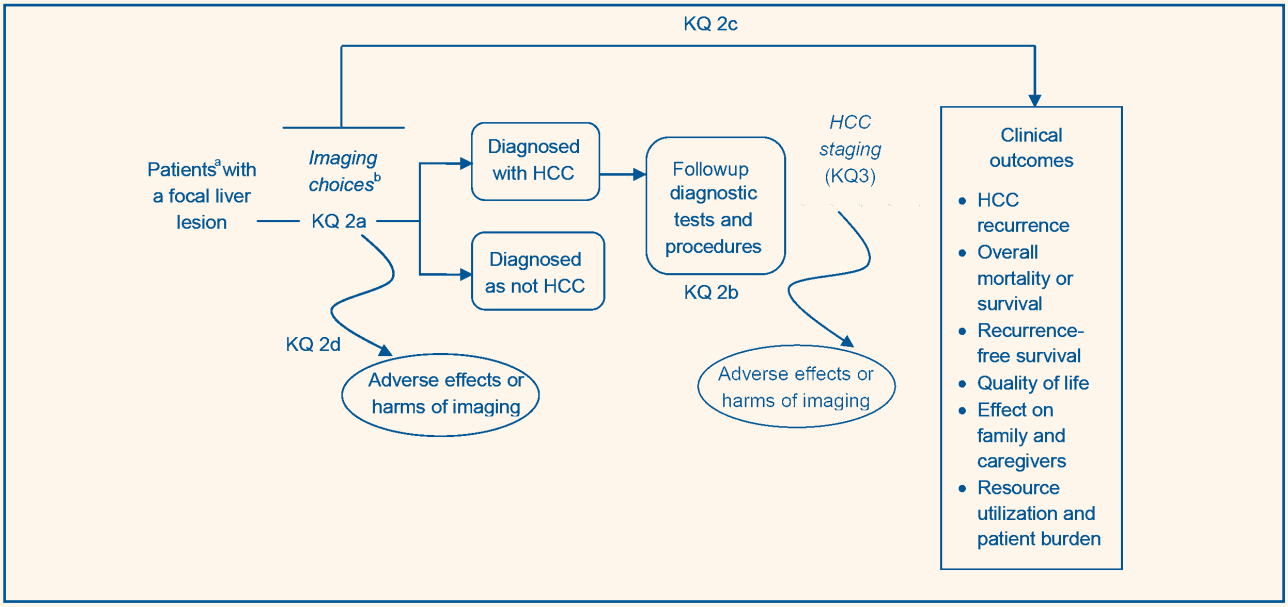
HCC = hepatocellular carcinoma; KQ = Key Question.

[a] Potential modifiers of test performance include patient (e.g., severity of liver disease, underlying cause of liver disease, body mass index, age, sex, race), tumor (e.g., tumor diameter, degree of differentiation, location), technical, and other factors (e.g., biomarker levels, setting).

[b] Imaging techniques are used singly, in combination, or in sequence with or without biomarkers used as modifiers.

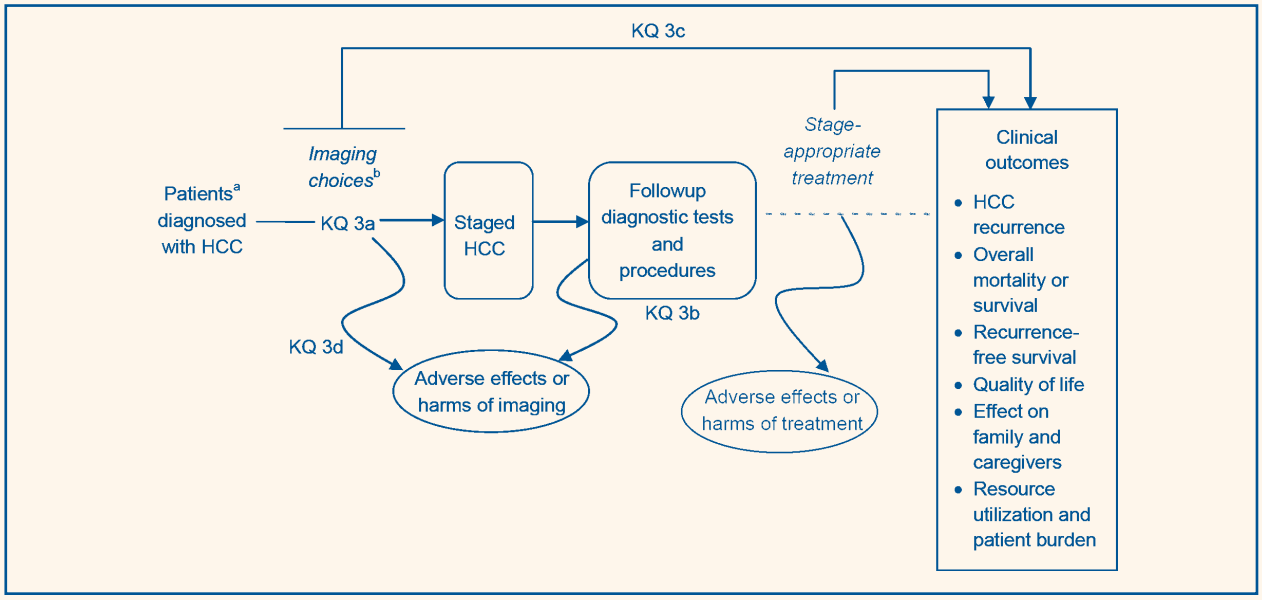
Note: Shaded figure elements illustrate the relationship of KQ 1 to KQ 2 and KQ 3.

Figure B. Analytic framework—evaluation of focal liver lesions (Key Question 2)



HCC = hepatocellular carcinoma; KQ = Key Question.
Note: Shaded elements show the relationship of KQ 2 to KQ 3.
[a] Potential modifiers of test performance include patient (e.g., severity of liver disease, underlying cause of liver disease, body mass index, age, sex, race), tumor (e.g., tumor diameter, degree of differentiation, location), technical, and other factors (e.g., biomarker levels, setting).
[b] Imaging techniques are used singly, in combination, or in sequence with or without biomarkers used as modifiers.

Figure C. Analytic framework—staging (Key Question 3)



[a] Potential modifiers of test performance include patient (e.g., severity of liver disease, underlying cause of liver disease, body mass index, age, sex, race), tumor (e.g., tumor diameter, degree of differentiation, location), technical, and other factors (e.g., biomarker levels, setting).
[b] Imaging techniques are used singly, in combination, or in sequence with or without biomarkers used as modifiers.
Note: Shaded elements show subsequent treatment that may follow detection (KQ 1), diagnosis (KQ 2), and staging (KQ 3).
HCC = hepatocellular carcinoma; KQ = Key Question.

Methods

The methods for this systematic review follow the methods suggested in the AHRQ Effective Health Care Program methods guides.[8,9]

Searching for the Evidence

For the primary literature, we searched Ovid MEDLINE®, Scopus, Evidence-Based Medicine Reviews (Ovid), the Cochrane Central Register of Controlled Trials, the Cochrane Database of Systematic Reviews, the Database of Abstracts of Reviews of Effects, and the Health Technology Assessment Database from 1998 to December 2013. We searched for unpublished studies in clinical trial registries (ClinicalTrials.gov, Current Controlled Trials, ClinicalStudyResults.org, and the World Health Organization International Clinical Trials Registry Platform), regulatory documents (U.S. Food and Drug Administration Medical Devices Registration and Listing), and individual product Web sites. Scientific information packets (SIPs) were solicited through the *Federal Register*.[10] We also searched the reference lists of relevant studies and previous systematic reviews for additional studies.

Study Selection

We developed criteria for inclusion and exclusion of studies based on the Key Questions and the populations, interventions, comparators, outcomes, timing, and setting (PICOTS) of interest. Titles and abstracts from all searches were reviewed for inclusion. Full-text articles were obtained for all articles identified as potentially meeting inclusion criteria. Papers were selected for inclusion in our review if they were about imaging for HCC with US (with or without contrast), CT with contrast, or MRI with contrast; were relevant to one or more Key Questions; met the predefined inclusion criteria; and reported original data.

We excluded studies that reported diagnostic accuracy of imaging for non-HCC malignant lesions; studies of nonspiral CT and MRI using machines ≤1.0 T, as these are considered outdated techniques;[11] studies that evaluated MRI with agents that are no longer produced commercially and are unavailable for clinical use; studies of CT arterial portography and CT hepatic angiography; studies published prior to 1998; studies in which imaging commenced prior to 1995, unless those studies reported use of imaging meeting minimum technical criteria; and studies of intraoperative US. We also excluded studies published only as conference abstracts, non–English-language articles, and studies of nonhuman subjects.

For studies of test performance (e.g., sensitivity, specificity, and likelihood ratios), we included studies that evaluated one or more imaging methods against a reference standard. Reference standards were histopathology (based on explanted liver or nonexplant histological specimen from surgery or percutaneous biopsy), imaging plus clinical followup (e.g., lesion growth), or some combination of these standards. We excluded studies in which the reference standard involved one of the imaging tests under evaluation and that did not perform clinical followup and studies that had no reference standard (i.e., reported the number of lesions identified with an imaging technique but did not evaluate accuracy against another reference technique).

To assess comparative effects of imaging on clinical outcomes (e.g., mortality, HCC recurrence, quality of life, and harms), we included randomized controlled trials that compared different imaging modalities or strategies. A systematic review funded by the Department of Veterans Affairs Evidence Synthesis Program on effects of screening for HCC on clinical outcomes is forthcoming and will include comparative observational studies.[12]

To assess comparative effects of imaging on intermediate outcomes related to clinical decisionmaking (e.g., subsequent diagnostic testing, treatments, or resource utilization), we included randomized trials and cohort studies that compared different imaging modalities or strategies.

Data Abstraction and Data Management

We extracted the following data from included studies into evidence tables using Excel spreadsheets: study design, year, setting, country, sample size, method of data collection (retrospective or prospective), eligibility criteria, population and clinical characteristics (including age, sex, race, underlying cause of liver disease, proportion of patients in sample with HCC, HCC lesion size, and proportion with cirrhosis), the number of imaging readers, criteria used for a positive test, and the reference standard used. We abstracted results for diagnostic accuracy, intermediate outcomes, and clinical outcomes, including results stratified according to patient, lesion, and imaging characteristics. Technical information for different imaging tests was abstracted.[11]

Assessment of Methodological Risk of Bias of Individual Studies

We assessed risk of bias (quality) for each study based on predefined criteria. Randomized trials and cohort studies were evaluated using criteria and methods developed

by the U.S. Preventive Services Task Force.[13] These criteria were applied in conjunction with the approach recommended in the Agency for Healthcare Research and Quality (AHRQ) "Methods Guide for Effectiveness and Comparative Effectiveness Reviews."[8] Studies of diagnostic test performance were assessed using the approach recommended in the AHRQ "Methods Guide for Medical Test Reviews,"[9] which is based on methods developed by the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) group.[14] Individual studies were rated as having "low," "moderate," or "high" risk of bias.

Data Synthesis

We performed meta-analyses on measures of test performance in order to help summarize data and obtain more precise estimates.[15] All quantitative analyses were conducted using SAS® 9.3 (SAS Institute Inc., Cary, NC). We pooled only studies that were clinically comparable and could provide a meaningful combined estimate (based on the variability among studies in design, patient population, imaging methods, and outcomes) and magnitude of effect size. We conducted separate analyses for each imaging modality, stratified according to the unit of analysis used (patients with HCC, HCC lesions, or liver segments with HCC) and analyzed studies of US with contrast separately from studies of US without contrast. For studies that used multiple readers, we averaged results across readers using the binomial specification of Proc NLMIXED on SAS.

We evaluated a number of potential sources of heterogeneity and modifiers of diagnostic accuracy. We performed analyses stratified according to the reference standard used and on domains related to risk of bias, aspects of study design (retrospective or prospective, use of a confidence rating scale), setting (based on country in which imaging was performed), and technical factors (such as scanner types, type of contrast or tracer used, use of recommended imaging phases, timing of delayed phase imaging, and section thickness). We also evaluated diagnostic accuracy in subgroups stratified according to HCC lesion size, degree of tumor differentiation, and tumor location, as well as patient characteristics such as severity of underlying liver disease, underlying cause of liver disease, and body mass index. Because of the effects of lesion size on estimates of diagnostic accuracy, subgroup analyses for each imaging modality were performed on the subgroup of studies that were not restricted to small (<2–3 cm) HCC lesions.

We performed separate analyses on the subset of studies that directly compared two or more imaging modalities or techniques in the same population against a common reference standard. Research indicates that results based on such direct comparisons differ from results based on noncomparative studies and may be better suited for evaluating comparative diagnostic test performance.[16]

We did not perform meta-analysis on staging accuracy and intermediate or clinical outcomes due to the small number of studies. Rather, we synthesized these studies qualitatively, using the methods described below for assessing the strength of evidence.

Grading the Strength of Evidence for Individual Comparisons and Outcomes

The strength of evidence for each Key Question was assessed by one researcher for each outcome described in the PICOTS using the approach described in the AHRQ "Methods Guide for Effectiveness and Comparative Effectiveness Reviews."[8] The strength of evidence pertains to the overall quality of each body of evidence and is based on the risk of bias (graded low, moderate, or high); the consistency of results between studies (graded consistent, inconsistent, or unknown/not applicable when only one study was available); the directness of the evidence linking the intervention and health outcomes (graded direct or indirect); and the precision of the estimate of effect, based on the number and size of studies and confidence intervals for the estimates (graded precise or imprecise). We did not assess studies of diagnostic test performance for publication bias using graphical or statistical methods because research indicates that such methods can be misleading. Rather, we searched for unpublished studies through searches of clinical trials registries and regulatory documents and by soliciting SIPs.

Assessing Applicability

We recorded factors important for understanding the applicability of studies, such as whether the publication adequately described the study population, the country in which the study was conducted, the prevalence of HCC in the patients who underwent imaging, the magnitude of differences in measures of diagnostic accuracy and clinical outcomes, and whether the imaging techniques were reasonably representative of standard practice.[17] We also recorded the funding source and role of the sponsor.

Results

The bulk of the available evidence addresses diagnostic accuracy of different imaging techniques for hepatocellular carcinoma. Few studies compared effects of different imaging modalities or strategies on clinical

decisionmaking and clinical outcomes, and almost no studies reported harms.

Results of Literature Searches

We reviewed titles and abstracts of the 4,846 citations identified by literature searches. Of these, 851 articles appeared to meet inclusion criteria and were selected for further full-text review. Following review at the full-text level, a total of 281studies met inclusion criteria.

We identified 274 studies that evaluated diagnostic accuracy of imaging tests. Of these, 70 evaluated US imaging, 134 evaluated CT, 129 evaluated MRI, and 32 evaluated PET; 28 studies evaluated more than one imaging modality. We rated 3 studies low risk of bias, 189 moderate risk of bias, and 89 high risk of bias. Almost all studies reported sensitivity, but only 130 reported specificity or provided data to calculate specificity. We found that 119 studies avoided use of a case-control design, 151 used blinded ascertainment, and 75 used a prospective design. More studies were conducted in Asia (182 studies) than in Australia, Canada, the United States, or Europe combined (92 studies). In 155 studies, imaging was conducted starting in or after 2003.

Data for outcomes other than measures of test performance were sparse. Seven studies reported comparative effects on clinical decisionmaking, three studies reported comparative clinical and patient-centered outcomes, and three studies reported harms associated with imaging for HCC.

Key Question 1. What is the comparative effectiveness of available imaging-based strategies, used singly or in sequence, for detection of HCC among individuals in surveillance and nonsurveillance settings?

Six studies evaluated diagnostic accuracy of imaging techniques for surveillance, and 182 studies reported diagnostic accuracy in nonsurveillance settings (e.g., imaging performed to assess detection rates in a series of patients undergoing treatment for HCC or patients with otherwise known prevalence of HCC prior to imaging). Four studies of PET evaluated accuracy specifically for identification of recurrent HCC. One randomized trial (rated high risk of bias) evaluated clinical outcomes associated with imaging-based surveillance versus no screening, and two trials evaluated clinical outcomes associated with different US surveillance intervals. No study compared effects of different imaging surveillance strategies on diagnostic thinking or clinical decisionmaking. Two studies reported harms associated with imaging for HCC. Tables A–F summarize the key findings and strength of evidence for these studies.

Key Question 2. What is the comparative effectiveness of imaging techniques, used singly, in combination, or in sequence, in diagnosing HCC among individuals in whom a focal liver lesion has been detected?

Fifty-four studies evaluated diagnostic accuracy of imaging tests in diagnosing HCC among individuals in whom an abnormal lesion has been detected, and 19 studies evaluated the accuracy of imaging tests for distinguishing HCC from another specific type of liver lesion. No study compared effects of different imaging modalities or strategies on diagnostic thinking or on clinical or patient-centered outcomes. One study reported harms. Tables G–L summarize the key findings and strength of evidence for these studies.

Key Question 3. What is the comparative effectiveness of imaging techniques, used singly, in combination, or in sequence, in staging HCC among patients diagnosed with HCC?

Six studies reported test performance of various imaging techniques for staging of patients with HCC based on TNM criteria. Ten studies reported test performance of PET for detection of metastatic disease. Seven studies reported effects of imaging on transplant decisions, and one study reported comparative effects of imaging on clinical and patient-centered outcomes. No study reported harms associated with imaging for HCC staging. Tables M–R summarize the key findings and strength of evidence for these studies.

Discussion

Key Findings and Strength of Evidence

The key findings of this review, including strength-of-evidence grades, are summarized in Tables A–R. The preponderance of evidence on imaging for HCC was in the area of diagnostic test performance. However, few studies evaluated test performance of imaging for HCC in true surveillance settings of patients at high risk for HCC, but without a prior diagnosis of HCC, undergoing periodic imaging. Among the limited evidence available in this setting, there was no clear difference between US without contrast and CT, based on across-study comparisons of sensitivity using either HCC lesions or patients with HCC as a unit of analysis. The strength of evidence is low for sensitivity. However, two studies that directly compared sensitivity of US without contrast and CT reported lower sensitivity with US for detection of patients with HCC.[18,19] The strength of evidence was also rated as low.

Many more studies evaluated test performance of imaging for HCC in populations of patients undergoing treatment such as liver transplantation, hepatic resection, or ablation therapy, or in series of patients previously diagnosed with HCC or with HCC and other liver conditions. Such studies were considered as part of Key Question 1 with studies of surveillance because they were not designed to further characterize previously identified HCC lesions (the focus of Key Question 2). Rather, their purpose was to evaluate test performance for lesion identification, therefore providing information that could potentially be extrapolated to surveillance. We analyzed these studies separately from studies conducted in true surveillance settings, given the differences in the reason for imaging and the populations evaluated, including a generally much higher prevalence of HCC, with some studies enrolling only patients with HCC. In these studies, sensitivity was lower for US without contrast than for CT or MRI, with a difference based on within-study (direct) comparisons that ranged from 0.11 to 0.22, using HCC lesions as the unit of analysis. This conclusion is graded moderate strength of evidence. MRI and CT performed similarly when patients with HCC were the unit of analysis, but sensitivity was higher for MRI than for CT when HCC lesions were the unit of analysis (pooled difference 0.09; 95% confidence interval, 0.07 to 12; moderate strength of evidence).

US with contrast did not perform better than US without contrast for identification of HCC[20,21] (low strength of evidence). This is probably related to the short duration in which microbubble contrast is present within the liver, so that it is not possible to perform a comprehensive contrast-enhanced examination of the liver.[22] Rather, the main use of US with contrast appears to be for evaluation of previously identified focal liver lesions.

For characterization of previously identified lesions, we found no clear differences in sensitivity between US with contrast, CT, and MRI (moderate strength of evidence). Although some evidence was available on the accuracy of imaging modalities for distinguishing between HCC and other (non-HCC) liver lesions, it was not possible to draw strong conclusions due to variability in the types of non-HCC lesions evaluated (regenerative nodules, dysplastic nodules, hypervascular pseudolesions, hemangiomas, etc.), small numbers of studies, and some inconsistency in findings.

Studies of patients with HCC were generally associated with somewhat higher sensitivity than studies that used HCC lesions as the unit of analysis. Studies that used explanted livers as the reference standard reported lower sensitivity than studies that used a nonexplant reference standard (moderate strength of evidence). Use of multiple reference standards poses a challenge to assessment of diagnostic accuracy.[23] Across imaging modalities, sensitivity was markedly lower for HCC lesions <2 cm versus those ≥2 cm (differences in sensitivity ranged from 0.30 to 0.39), and further declined for lesions <10 mm in diameter (moderate strength of evidence). Evidence also consistently indicated substantially lower sensitivity for well-differentiated lesions than moderately or poorly differentiated lesions (low strength of evidence).

Evidence on the effects of other patient, tumor, and technical factors on test performance was more limited (low strength of evidence). For US, there was no clear effect of use of Doppler, lesion depth, or body mass index on test performance. For CT, some evidence indicated higher sensitivity for studies that used a contrast rate of ≥3 ml/s than those with a contrast rate <3 ml/s, and higher sensitivity for studies that used delayed phase imaging. For MRI, hepatic-specific contrast agents were associated with slightly higher sensitivity than nonspecific contrast agents, but there were no clear effects of magnetic field strength (3.0 vs. 1.5 T), use of delayed phase imaging, timing of delayed phase imaging (≥120 seconds after administration of contrast or <120 s), section thickness (≤5 mm vs. >5 mm), or use of diffusion-weighted imaging. For identification of intrahepatic HCC lesions, limited evidence found PET with ^{11}C-acetate and other alternative tracers such as ^{18}F-fluorocholine and ^{18}F-fluorothymidine associated with substantially higher sensitivity than 18F-fluorodeoxyglucose (FDG) PET. Sensitivity of FDG PET was lower than sensitivity of FDG PET/CT.

The limited available evidence suggests that using multiple imaging tests and defining a positive test as typical imaging findings on at least one imaging modality increases sensitivity without substantively reducing specificity (moderate strength of evidence).

Conclusions were generally robust on sensitivity and stratified analyses based on study factors such as setting (Asia vs. United States or Europe), prospective collection of data, interpretation of imaging findings blinded to results of the reference standard, avoidance of case-control design, and overall risk of bias.

Across analyses, specificity was generally high, with most pooled estimates around 0.85 or higher and few clear differences between imaging modalities. However, many studies did not report specificity and pooled estimates of specificity were frequently imprecise, precluding strong conclusions regarding comparative test performance. Since likelihood ratios are sensitive to small changes in estimates

when the specificity is high, it was also difficult to draw strong conclusions regarding comparative diagnostic test performance based on differences in positive or negative likelihood ratios. Most likelihood ratio estimates fell into or near the "moderately useful" range (positive likelihood ratio of 5–10 and negative likelihood ratio of 0.1–0.2), with the exception of FDG PET for identification of intrahepatic HCC lesions, which was associated with a negative likelihood ratio of 0.50.

Evidence regarding the accuracy of imaging modalities for staging was primarily limited to CT. Most studies addressed accuracy of CT, with 28 to 58 percent correctly staged based on TNM criteria and somewhat more understaging (25% to 52%) than overstaging (2% to 27%) (moderate strength of evidence). Studies on the accuracy of imaging for identifying metastatic HCC disease were primarily limited to FDG PET or PET/CT, with a pooled sensitivity of 0.82 to 0.85 (low strength of evidence).

Evidence on the comparative effectiveness of imaging for HCC on diagnostic thinking, use of subsequent procedures, or resource utilization was extremely limited. In studies that compared the accuracy of transplant decisions based on CT against primarily explanted livers as the reference standard, the proportion correctly assessed for transplant eligibility based on Milan criteria ranged from 40 to 96 percent (moderate strength of evidence). Evidence on the effects of surveillance with imaging versus no surveillance on clinical outcomes was limited to a single randomized trial[24] (low strength of evidence).

Evidence on comparative harms associated with imaging was also extremely limited, with no study measuring downstream harms related to false-positive tests or subsequent workup, or potential harms related to labeling or psychological effects. A handful of studies reported low rates of serious direct harms (e.g., allergic reactions) associated with imaging. However, evidence on administration of contrast for radiological procedures in general also suggests a low rate of serious adverse events. No study on US with contrast reported harms. Although PET and CT are associated with risk of radiation exposure, no study of imaging for HCC was designed to evaluate potential long-term clinical outcomes associated with radiation exposure.

Table A. Summary of evidence for Key Question 1.a (detection): test performance

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Surveillance settings *Unit of analysis: patients with HCC*	US without contrast	Sensitivity: Low Specificity: Low	Sensitivity was 0.78 (95% CI, 0.60 to 0.89; 4 studies) and specificity 0.89 (95% CI, 0.80 to 0.94; 3 studies), for an LR+ of 6.8 (95% CI, 4.2 to 11) and LR- of 0.25 (95% CI, 0.13 to -0.46).
Surveillance settings *Unit of analysis: patients with HCC*	CT	Sensitivity: Low Specificity: Low	Sensitivity was 0.84 (95% CI, 0.59 to 0.95; 2 studies) and specificity 0.999 (95% CI, 0.86 to 0.99; 2 studies), for an LR+ of 60 (95% CI, 5.9 to 622) and LR- of 0.16 (95% CI, 0.06 to 0.47).
Surveillance settings *Unit of analysis: patients with HCC*	MRI or PET	Insufficient	No evidence
Surveillance settings *Unit of analysis: HCC lesions*	US without contrast	Sensitivity: Low Specificity: Insufficient	Sensitivity was 0.60 (95% CI, 0.24 to 0.87; 1 study); specificity was not reported.

Table A. Summary of evidence for Key Question 1.a (detection): test performance (continued)

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Surveillance settings *Unit of analysis: HCC lesions*	CT	Sensitivity: Low Specificity: Insufficient	Sensitivity was 0.62 (95% CI, 0.46 to 0.76; 1 study).
Surveillance settings *Unit of analysis: HCC lesions*	MRI or PET	Insufficient	No evidence
Nonsurveillance settings *Unit of analysis: patients with HCC*	US without contrast	Sensitivity: Low Specificity: Low	Sensitivity was 0.73 (95% CI, 0.46 to 0.90; 8 studies) and specificity 0.93 (95% CI, 0.85 to 0.97; 6 studies), for an LR+ of 11 (95% CI, 5.4 to 21) and LR- of 0.29 (95% CI, 0.13 to 0.65).
Nonsurveillance settings *Unit of analysis: patients with HCC*	CT	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.83 (95% CI, 0.75 to 0.89; 16 studies) and specificity 0.92 (95% CI, 0.86 to 0.96; 11 studies), for an LR+ of 11 (95% CI, 5.6 to 20) and LR- of 0.19 (95% CI, 0.12 to 0.28).
Nonsurveillance settings *Unit of analysis: patients with HCC*	MRI	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.85 (95% CI, 0.76 to 0.91; 10 studies) and specificity 0.90 (95% CI, 0.81 to 0.94; 8 studies), for an LR+ of 8.1 (95% CI, 4.3 to 15) and LR- of 0.17 (95% CI, 0.10 to 0.28).
Nonsurveillance settings *Unit of analysis: patients with HCC*	PET	Sensitivity: Moderate Specificity: Low	For FDG PET, sensitivity was 0.52 (95% CI, 0.39 to 0.66; 15 studies) and specificity was 0.95 (95% CI, 0.82 to 0.99; 5 studies), for an LR+ of 11 (95% CI, 2.6 to 49) and LR- of 0.50 (95% CI, 0.37 to 0.68). For 11C-acetate PET or PET/CT, sensitivity was 0.85 (95% CI, 0.67 to 0.94; 4 studies); specificity was not reported.
Nonsurveillance settings *Unit of analysis: HCC lesions*	US without contrast	Sensitivity: Moderate Specificity: Low	Sensitivity was 0.59 (95% CI, 0.42 to 0.74; 11 studies). Only 2 studies reported specificity, with inconsistent results (0.63; 95% CI, 0.53 to 0.73, and 0.95; 95% CI, 0.85 to 0.99).
Nonsurveillance settings *Unit of analysis: HCC lesions*	US with contrast	Sensitivity: Low Specificity: Insufficient	Sensitivity was 0.73 (95% CI, 0.52 to 0.87; 8 studies). No study evaluated specificity.
Nonsurveillance settings *Unit of analysis: HCC lesions*	CT	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.77 (95% CI, 0.72 to 0.80; 79 studies) and specificity 0.89 (95% CI, 0.84 to 0.93; 21 studies), for an LR+ of 7.1 (95% CI, 4.7 to 11) and LR- of 0.26 (95% CI, 0.22 to 0.31).
Nonsurveillance settings *Unit of analysis: HCC lesions*	MRI	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.82 (95% CI, 0.79 to 0.85; 75 studies) and specificity 0.87 (95% CI, 0.77 to -0.93; 16 studies), for an LR+ of 6.4 (95% CI, 3.5 to 12) and LR- of 0.20 (95% CI, 0.16 to 0.25).

Table A. Summary of evidence for Key Question 1.a (detection): test performance (continued)

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Nonsurveillance settings *Unit of analysis: HCC lesions*	PET	Sensitivity: Moderate Specificity: Low	For FDG PET, sensitivity was 0.53 (95% CI, 0.41 to 0.65; 5 studies) and specificity 0.91 (95% CI, 0.76 to 0.98; 1 study). For 11C-acetate PET, sensitivity was 0.78 (95% CI, 0.61 to 0.89; 4 studies); specificity was not reported.
Direct (within-study) comparisons of imaging modalities *Unit of analysis: patients with HCC*	US without contrast vs. CT	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.68 (95% CI, 0.54 to 0.80) vs. 0.80 (95% CI, 0.68 to 0.88), for a difference of -0.12 (95% CI, -0.20 to -0.03), based on 6 studies. Two studies were performed in surveillance settings. (Low strength of evidence for sensitivity and specificity.)
Direct (within-study) comparisons of imaging modalities *Unit of analysis: patients with HCC*	US without contrast vs. MRI	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.61 (95% CI, 0.48 to 0.74) vs. 0.81 (95% CI, 0.69 to 0.89), for a difference of -0.19 (95% CI, -0.30 to -0.08), based on 3 studies, none of which were performed in surveillance settings.
Direct (within-study) comparisons of imaging modalities *Unit of analysis: patients with HCC*	MRI vs. CT	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.88 (95% CI, 0.53 to 0.98) vs. 0.82 (95% CI, 0.41 to 0.97), for a difference of 0.06 (95% CI, -0.05 to 0.17), based on 4 studies, none of which were performed in surveillance settings.
Direct (within-study) comparisons of imaging modalities *Unit of analysis: HCC lesions*	US without contrast vs. CT	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.55 (95% CI, 0.43 to 0.66) vs. 0.66 (95% CI, 0.54 to 0.76), for a difference of -0.11 (95% CI, -0.18 to -0.04), based on 3 studies, none of which were performed in surveillance settings.
Direct (within-study) comparisons of imaging modalities *Unit of analysis: HCC lesions*	US without contrast vs. MRI	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.57 (95% CI, 0.42 to 0.71) vs. 0.79 (95% CI, 0.67 to 0.88), for a difference of -0.22 (95% CI, -0.31 to 0.14), based on 3 studies, none of which were performed in surveillance settings.

Table A. Summary of evidence for Key Question 1.a (detection): test performance (continued)

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Direct (within-study) comparisons of imaging modalities *Unit of analysis: HCC lesions*	US with contrast vs. CT	Sensitivity: Moderate Specificity: Insufficient	Sensitivity was 0.51 (95% CI, 0.29 to 0.74) vs. 0.61 (95% CI, 0.38 to 0.81), for a difference of -0.10 (95% CI, -0.20 to -0.00), based on 4 studies, none of which were performed in surveillance settings.
Direct (within-study) comparisons of imaging modalities *Unit of analysis: HCC lesions*	US with contrast vs. MRI	Sensitivity: Moderate Specificity: Insufficient	Sensitivity was 0.65 (95% CI, 0.41 to 0.84) vs. 0.73 (95% CI, 0.50 to 0.88), for a difference of -0.08 (95% CI, -0.19 to 0.02), based on 3 studies, none of which were performed in surveillance settings.
Direct (within-study) comparisons of imaging modalities *Unit of analysis: HCC lesions*	MRI vs. CT	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.81 (95% CI, 0.76 to 0.84) vs. 0.71 (95% CI, 0.66 to 0.76), for a difference of 0.09 (95% CI, 0.07 to 0.12), based on 31 studies, none of which were performed in surveillance settings. Findings were similar when studies were stratified according to use of nonhepatic-specific or hepatic-specific contrast and when the analysis was restricted to HCC lesions <2–3 cm. For HCC lesions <2–3 cm, the difference in sensitivity was greater for studies of hepatic-specific MRI contrast (0.23; 95% CI, 0.17 to 0.29; 12 studies) than for studies of nonhepatic-specific MRI contrast (0.06; 95% CI, -0.01 to 0.13; 6 studies).
Multiple imaging modalities	Various combinations	Sensitivity: Insufficient Specificity: Insufficient	One study found sensitivity of imaging with various combinations of 2 imaging modalities was similar or lower than with single-modality imaging, based on concordant positive findings on 2 imaging modalities. The other study reported higher sensitivity with multiple imaging modalities than with single-modality imaging, but criteria for positive results based on multiple imaging modalities were unclear.

CI = confidence interval; CT = computed tomography; FDG = ^{18}F-fluorodeoxyglucose; HCC = hepatocellular carcinoma; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; MRI = magnetic resonance imaging; PET = positron emission tomography; US = ultrasound.

Table B. Summary of evidence for Key Question 1.a.i (detection): effects of reference standard on test performance (based on HCC lesions as the unit of analysis)

Imaging Modality or Comparison	Strength of Evidence	Summary
US without contrast	Sensitivity: Moderate Specificity: Insufficient	Using HCC lesions as the unit of analysis, sensitivity was 0.34 (95% CI, 0.22 to 0.47) in 5 studies that used explanted liver as the reference standard and ranged from 0.72 to 0.75 in studies that used other reference standards.
US with contrast	Sensitivity: Low Specificity: Insufficient	No study using HCC lesions as the unit of analysis used an explanted liver reference standard. Sensitivity was 0.58 (95% CI, 0.39 to 0.75) using a nonexplant histopathological reference standard and 0.98 (95% CI, 0.88 to 0.997) using a mixed reference standard.
CT	Sensitivity: Moderate Specificity: Moderate	Using HCC lesions as the unit of analysis, sensitivity was 0.67 (95% CI, 0.59 to 0.75) in 23 studies that used explanted liver as the reference standard and ranged from 0.65 to 0.86 in studies that used other reference standards.
MRI	Sensitivity: Moderate Specificity: Moderate	Using HCC lesions as the unit of analysis, sensitivity was 0.69 (95% CI, 0.59 to 0.77) in 15 studies that used explanted liver as the reference standard and ranged from 0.85 to 0.88 in studies that used a nonexplant histopathological reference standard or mixed reference standard; only 3 studies evaluated an imaging/clinical reference standard (sensitivity, 0.65; 95% CI, 0.43 to 0.83).
PET	Sensitivity: Low Specificity: Insufficient	No study of FDG PET used an explanted liver reference standard. Four of the 5 studies that used HCC lesions as the unit of analysis used a nonexplant histological reference standard (sensitivity, 0.49; 95% CI, 0.37 to 0.61).

CI = confidence interval; CT = computed tomography; FDG = ^{18}F-fluorodeoxyglucose; HCC = hepatocellular carcinoma; MRI = magnetic resonance imaging; PET = positron emission tomography; US = ultrasound.

Table C. Summary of evidence for Key Question 1.a.ii (detection): effects of patient, tumor, technical, and other factors on test performance

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Lesion size	US without contrast	Sensitivity: Moderate Specificity: Low	Sensitivity was 0.82 (95% CI, 0.68 to 0.91) for lesions ≥2 cm and 0.34 (95% CI, 0.19 to 0.53) for lesions <2 cm, for a difference of 0.48 (95% CI, 0.39 to 0.57). Sensitivity was 0.09 (95% CI, 0.02 to 0.29; 4 studies) for lesions <10 mm, to 0.50 (95% CI, 0.23 to 0.78; 4 studies) for lesions 10–20 mm and 0.88 (95% CI, 0.66 to 0.96; 4 studies) for lesions >20 mm, for a difference of 0.37 (95% CI, 0.18 to 0.57) for lesions >20 mm vs. 10–20 mm and 0.41 (95% CI, 0.19 to 0.63) for lesions 10–20 mm vs. <10 mm.
Lesion size	US with contrast	Sensitivity: Low Specificity: Low	Sensitivity was 0.94 (95% CI, 0.83 to 0.98) for lesions ≥>2 cm and 0.77 (95% CI, 0.53 to 0.91) for lesions <2 cm, for a difference of 0.17 (95% CI, 0.03 to 0.32), based on 5 studies. Three studies found sensitivity of 0.64 (95% CI, 0.33 to 0.87) for lesions 10–20 mm and 0.91 (95% CI, 0.71 to 0.98) for lesions >20 mm, for a difference of 0.26 (95% CI, 0.04 to 0.48).

Table C. Summary of evidence for Key Question 1.a.ii (detection): effects of patient, tumor, technical, and other factors on test performance (continued)

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Lesion size	CT	Sensitivity: Moderate Specificity: Low	Sensitivity was 0.94 (95% CI, 0.92 to 0.95) for lesions ≥2 cm and 0.63 (95% CI, 0.57 to 0.69) for lesions <2 cm, for an absolute difference in sensitivity of 0.31 (95% CI, 0.26 to 0.36), based on 34 studies. Sensitivity was 0.32 (95% CI, 0.25 to 0.41; 21 studies) for lesions <10 mm, 0.74 (95% CI, 0.67 to 0.80; 23 studies) for lesions 10–20 mm, and 0.95 (95% CI, 0.92 to 0.97; 20 studies), for a difference of 0.21 (95% CI, 0.15 to 0.26) for lesions >20 vs. 10–20 mm and 0.42 (95% CI, 0.36 to 0.48) for lesions 10–20 vs. <10 mm.
Lesion size	MRI	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.96 (95% CI, 0.93 to 0.97) for lesions ≥2 cm and 0.66 (95% CI, 0.58 to 0.74) for lesions <2 cm, for an absolute difference in sensitivity of 0.29 (95% CI, 0.23 to 0.36), based on 29 studies. Sensitivity was 0.45 (95% CI, 0.34 to 0.56; 20 studies) for lesions <10 mm, 0.78 (95% CI, 0.69 to 0.85; 21 studies) for lesions 10–20 mm, and 0.97 (95% CI, 0.94 to 0.98, 14 studies) for lesions >20 mm (95% CI, 0.94 to 0.98; 18 studies), for a difference of 0.19 (95% CI, 0.12 to 0.26) for >20 vs. 10–20 mm and 0.33 (95% CI, 0.26 to 0.40) for 10–20 vs. <10 mm.
Lesion size	PET	Sensitivity: Low Specificity: Insufficient	For FDG PET, sensitivity was consistently higher for larger lesions, based on 5 studies. Data were not pooled due to differences in the tumor size categories evaluated. Two studies of 11C-acetate PET found inconsistent effects of lesion size on sensitivity.
Degree of tumor differentiation	US with contrast	Sensitivity: Low Specificity: Insufficient	Sensitivity was 0.83 (95% CI, 0.55 to 0.95) for moderately or poorly differentiated HCC lesions and 0.43 (95% CI, 0.15 to 0.76) for well differentiated lesions, for an absolute difference in sensitivity of 0.40 (95% CI, 0.17 to 0.64), based on 3 studies.
Degree of tumor differentiation	CT	Sensitivity: Low Specificity: Insufficient	Sensitivity was 0.82 (95% CI, 0.66 to 0.91) for moderately or poorly differentiated HCC lesions and 0.50 (95% CI, 0.29 to 0.70) for well differentiated lesions, for an absolute difference in sensitivity of 0.32 (95% CI, 0.19 to 0.45), based on 5 studies.
Degree of tumor differentiation	MRI	Sensitivity: Low Specificity: Insufficient	Sensitivity was 0.68 (95% CI, 0.44 to 0.86) for moderately or poorly differentiated HCC lesions and 0.37 (95% CI, 0.17 to 0.62) for well differentiated lesions, for an absolute difference in sensitivity of 0.31 (95% CI, 0.13 to 0.49), based on 3 studies.
Degree of tumor differentiation	PET	Sensitivity: Low Specificity: Insufficient	For FDG PET, sensitivity was 0.72 (95% CI, 0.59 to 0.83) for moderately or poorly differentiated HCC lesions and 0.39 (95% CI, 0.26 to 0.55) for well differentiated lesions, for an absolute difference in sensitivity of 0.33 (95% CI, 0.20 to 0.46), based on 6 studies. In 3 studies of 11C-acetate PET and 1 study of 18F-fluorochorine PET, sensitivity for more well differentiated lesions was not lower than for more poorly differentiated lesions.
Other factors	US	Low	In 2 studies that directly compared US with vs. without contrast, there was no clear difference in sensitivity (-0.04; 95% CI, -0.11 to 0.04). One study that directly compared use of Doppler vs. no Doppler showed no clear effect on estimates of sensitivity. Lesion depth and body mass index had no effect on estimates of sensitivity.

Table C. Summary of evidence for Key Question 1.a.ii (detection): effects of patient, tumor, technical, and other factors on test performance (continued)

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Other factors	CT	Low	Using patients with HCC as the unit of analysis, studies with a contrast rate ≥3 ml/s reported a higher sensitivity (0.87; 95% CI, 0.77 to 0.93; 8 studies) than studies with a contrast rate <3 ml/s (0.71; 95% CI, 0.50 to -0.85; 4 studies). Studies with delayed phase imaging reported somewhat higher sensitivity (0.89; 95% CI, 0.81 to 0.94; 7 studies) than studies without delayed phase imaging (0.74; 95% CI, 0.66 to 0.87; 7 studies). However, neither of these technical parameters had clear effects in studies that used HCC lesions as the unit of analysis.
Other factors	MRI	Low	There were no clear differences in estimates of diagnostic accuracy when studies were stratified according to MRI scanner type (1.5 vs. 3.0 T), imaging phases evaluated (with or without delayed phase imaging), timing of delayed phase imaging (≥120 seconds vs. <120 seconds), section thickness (≤5 mm for enhanced images vs. >5 mm), or use of diffusion-weighted imaging. In studies that directly compared diagnostic accuracy with different types of contrast, hepatic-specific contrast agents were associated with slightly higher sensitivity than nonhepatic-specific contrast agents (0.83; 95% CI, 0.75 to 0.90, vs. 0.74; 95% CI, 0.62 to 0.83; difference 0.10; 95% CI, 0.04 to 0.15; 6 studies).
Other factors	PET	Low	FDG PET was associated with lower sensitivity that 11C-acetate PET when either patients (0.58 vs. 0.81, for a difference of -0.23; 95% CI, -0.34 to -0.13; 3 studies) or HCC lesions (0.52 vs. 0.79, for a difference of -0.27; 95% CI, -0.36 to -0.17; 3 studies) were the unit of analysis. FDG PET was also associated with lower sensitivity that dual tracer PET with FDG and 11C-acetate or 18F-choline PET, but evidence was limited to 1 or 2 studies for each of these comparisons. Using patients as the unit of analysis, sensitivity of FDG PET (0.39; 95% CI, 0.24 to 0.56; 8 studies) was lower than sensitivity of FDG PET/CT (0.65; 95% CI, 0.50 to 0.78; 7 studies).

CI = confidence interval; CT = computed tomography; FDG = ^{18}F-fluorodeoxyglucose; HCC = hepatocellular carcinoma; MRI = magnetic resonance imaging; PET = positron emission tomography; US = ultrasound.

Table D. Summary of evidence for Key Question 1.b (detection): clinical decisionmaking

Imaging Modality or Comparison	Strength of Evidence	Summary
Effects of different imaging modalities or strategies on clinical decisionmaking	Low	One randomized controlled trial (n = 163) found no clear differences between surveillance with US without contrast vs. CT in HCC detection rates, subsequent imaging, or cost per HCC detected.

CT = computed tomography; HCC = hepatocellular carcinoma; US = ultrasound.

Table E. Summary of evidence for Key Question 1.c (detection): clinical and patient-centered outcomes

Imaging Modality or Comparison	Strength of Evidence	Summary
US plus serum AFP	Low	One cluster randomized controlled trial (n = 18,816) conducted in China found screening every 6 months with noncontrast US plus serum AFP vs. no screening in persons 35 to 79 years of age (mean, 42 years) with HBV infection or chronic hepatitis without HBV infection to be associated with lower risk of HCC-related mortality (32 vs. 54 deaths; rate ratio, 0.63; 95% CI, 0.41 to 0.98) at 5-year followup, but was rated high risk of bias due to multiple methodological shortcomings.
US screening at different intervals, mortality	Moderate	Two trials (n = 2,022) found no clear differences in mortality with US screening at 4- vs. 12-month intervals, or at 3- vs. 6-month intervals. One trial (n = 163) found no difference in HCC mortality between surveillance with US without contrast vs. CT, but was underpowered to detect differences.

AFP = alpha-fetoprotein; CI = confidence interval; CT = computed tomography; HBV = hepatitis B virus; HCC = hepatocellular carcinoma; US = ultrasound.

Table F. Summary of evidence for Key Question 1.d (detection): harms

Imaging Modality or Comparison	Strength of Evidence	Summary
MRI, CT, US	Insufficient	One study reported no serious adverse events associated with administration of gadoxetic acid for MRI, and 1 study reported no clear differences in adverse events between CT with contrast at 3 ml/s vs. 5 ml/s. No study reported rates of adverse events associated with use of microbubble contrast agents in US, and harms were not reported in randomized trials of screening with imaging.

CT = computed tomography; MRI = magnetic resonance imaging; US = ultrasound.

Table G. Summary of evidence for Key Question 2.a (evaluation of focal liver lesions): test performance

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Evaluation of focal liver lesion *Unit of analysis: patients with HCC*	US with contrast	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.87 (95% CI, 0.79 to 0.92; 12 studies) and specificity 0.91 (95% CI, 0.83 to 0.95; 8 studies), for an LR+ of 9.6 (95% CI, 5.1 to 18) and LR- of 0.14 (95% CI, 0.09 to 0.23).
Evaluation of focal liver lesion *Unit of analysis: patients with HCC*	US without contrast	Sensitivity: Low Specificity: Insufficient	Sensitivity was 0.78 (95% CI, 0.69 to 0.86) in 1 study; specificity was not reported.

Table G. Summary of evidence for Key Question 2.a (evaluation of focal liver lesions): test performance (continued)

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Evaluation of focal liver lesion *Unit of analysis: patients with HCC*	CT	Sensitivity: Moderate Specificity: Low	Sensitivity was 0.86 (95% CI, 0.75 to 0.92; 8 studies) and specificity 0.88 (95% CI, 0.76 to 0.95; 5 studies), for an LR+ of 7.4 (95% CI, 3.3 to 17) and LR- of 0.16 (95% CI, 0.09 to 0.30).
Evaluation of focal liver lesion *Unit of analysis: patients with HCC*	MRI	Sensitivity: Low Specificity: Low	Sensitivity was 0.77 (95% CI, 0.66 to 0.84; 4 studies) and specificity was 0.81 (95% CI, 0.52 to 0.94; 4 studies), for an LR+ of 4.0 (95% CI, 1.4 to 12) and LR- of 0.29 (95% CI, 0.21 to 0.39).
Evaluation of focal liver lesion *Unit of analysis: HCC lesions*	US with contrast	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.87 (95% CI, 0.80 to 0.92; 21 studies) and specificity 0.91 (95% CI, 0.85 to 0.95; 10 studies) for an LR+ of 9.8 (95% CI, 5.7 to 17) and LR- of 0.14 (95% CI, 0.09 to 0.23).
Evaluation of focal liver lesion *Unit of analysis: HCC lesions*	CT	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.79 (95% CI, 0.67 to 0.87; 13 studies) and specificity 0.90 (95% CI, 0.37 to 0.99; 6 studies), for an LR+ of 7.7 (95% CI, 0.71 to 84) and LR- of 0.24 (95% CI, 0.15 to 0.38).
Evaluation of focal liver lesion *Unit of analysis: HCC lesions*	MRI	Sensitivity: Moderate Specificity: Moderate	Sensitivity was 0.81 (95% CI, 0.72 to 0.87; 14 studies) and specificity 0.93 (95% CI, 0.80 to 0.98; 11 studies), for an LR+ of 12 (95% CI, 3.8 to 39) and LR- of 0.21 (95% CI, 0.15 to 0.30).
Evaluation of focal liver lesion *Unit of analysis: HCC lesions*	PET	Sensitivity: Low Specificity: Low	Sensitivity was 0.56 to 0.57 and specificity 1.0 in 2 studies of FDG PET.
For distinguishing HCC lesions from non-HCC hepatic lesions	US with contrast	Sensitivity: Low Specificity: Low	One study found US with sulfur hexafluoride contrast associated with a sensitivity of 0.94 (62/66) and a specificity of 0.68 (23/34) for distinguishing hypervascular HCC from focal nodular hyperplasia using quantitative methods, and 1 study found US with perflubutane contrast associated with a sensitivity of 0.59 (32/54) and specificity of 1.0 (13/13) for distinguishing small (<3 cm) well differentiated HCC lesions from regenerative nodules.
For distinguishing HCC lesions from non-HCC hepatic lesions	CT	Sensitivity: Low Specificity: Low	Five studies evaluated accuracy of CT for distinguishing HCC from non-HCC lesions, but the non-HCC lesions varied in the studies, precluding strong conclusions.
For distinguishing HCC lesions from non-HCC hepatic lesions	MRI	Sensitivity: Moderate Specificity: Moderate	Four studies reported inconsistent results for distinguishing small (<2 to 3 cm) hypervascular HCC lesions from hypervascular pseudolesions, with sensitivity 0.47 and 0.52 in 2 studies, and 0.91 and 0.92 in the other 2. Specificity was 0.93 or higher in all 4 studies. Eight other studies evaluated accuracy of MRI for distinguishing HCC from non-HCC lesions, but the non-HCC hepatic lesions varied in the studies.

Table G. Summary of evidence for Key Question 2.a (evaluation of focal liver lesions): test performance (continued)

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Direct (within-study) comparisons of imaging modalities *Unit of analysis: patients with HCC*	US without contrast vs. CT	Sensitivity: Low Specificity: Insufficient	Sensitivity was 0.78 (95% CI, 0.70 to 0.85) vs. 0.89 (95% CI, 0.84 to 0.95), for a difference of -0.12 (95% CI, -0.21 to -0.02), based on 1 study.
Direct (within-study) comparisons of imaging modalities *Unit of analysis: patients with HCC*	US with contrast vs. CT	Sensitivity: Moderate Specificity: Low	Sensitivity was 0.91 (95% CI, 0.85 to 0.94) vs. 0.88 (95% CI, 0.81 to 0.92), for a difference of 0.03 (95% CI, -0.02 to 0.08), based on 5 studies.
Direct (within-study) comparisons of imaging modalities *Unit of analysis: patients with HCC*	MRI vs. CT	Sensitivity: Low Specificity: Low	Sensitivity was 0.81 (95% CI, 0.70 to 0.92) vs. 0.74 (95% CI, 0.62 to 0.87), for a difference of 0.06 (-0.10 to 0.23), based on 1 study.
Direct (within-study) comparisons of imaging modalities *Unit of analysis: HCC lesions*	US with contrast vs. CT	Sensitivity: Moderate Specificity: Insufficient	Sensitivity was 0.92 (95% CI, 0.88 to 0.96) vs. 0.89 (95% CI, 0.83 to 0.93), for a difference of 0.04 (95% CI, -0.02 to 0.09), based on 4 studies.
Direct (within-study) comparisons of imaging modalities *Unit of analysis: HCC lesions*	US with contrast vs. MRI	Sensitivity: Low Specificity: Low	Sensitivity was 0.79 (95% CI, 0.65 to 0.94) vs. 0.83 (95% CI, 0.69 to 0.97), for a difference of -0.03 (95% CI, -0.24 to 0.17), based on 1 study.

Table G. Summary of evidence for Key Question 2.a (evaluation of focal liver lesions): test performance (continued)

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Direct (within-study) comparisons of imaging modalities *Unit of analysis: HCC lesions*	MRI vs. CT	Sensitivity: Low Specificity: Low	One study found MRI associated with higher sensitivity (0.84; 95% CI, 0.76 to 0.92 vs. 0.62; 95% CI, 0.52 to 0.72, for a difference of 0.22; 95% CI, 0.09 to 0.35) but lower specificity (0.36; 95% CI, 0.20 to 0.52 vs. 0.72; 95% CI, 0.58 to 0.87, for a difference of -0.36; 95% CI, -0.58 to 0.15) than CT.
Multiple imaging modalities	Various combinations	Moderate	In 4 studies in which positive results with multiple modality imaging were defined as concordant typical findings for HCC on 2 imaging modalities, sensitivity was lower than with a single modality (difference in sensitivity ranged from 0.09 to 0.27), with no clear difference in specificity. In 3 studies in which positive results with multiple modality imaging were defined as typical findings for HCC on at least 1 of the imaging techniques, sensitivity was higher than with a single modality (increase in sensitivity ranged from 0.09 to 0.25), with no clear difference in specificity. One study found that a sequential imaging strategy in which a second imaging test was performed only for indeterminate results on initial CT increased sensitivity for HCC from 0.53 to 0.74 to 0.79.

CI = confidence interval; CT = computed tomography; FDG = ^{18}F-fluorodeoxyglucose; HCC = hepatocellular carcinoma; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; MRI = magnetic resonance imaging; PET = positron emission tomography; US = ultrasound.

Table H. Summary of evidence for Key Question 2.a.i (evaluation of focal liver lesions): effects of reference standard on test performance (based on HCC lesions as the unit of analysis)

Imaging Modality or Comparison	Strength of Evidence	Summary
All	Sensitivity: Moderate Specificity: Moderate	No study used explanted liver as the reference standard. There were no clear differences across imaging modalities in estimates of diagnostic accuracy in analyses stratified by use of different nonexplant reference standards.

HCC = hepatocellular carcinoma.

Table I. Summary of evidence for Key Question 2.a.ii (evaluation of focal liver lesions): effects of patient, tumor, technical, and other factors on test performance

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Other factors	US	Low	In 2 studies that directly compared US with vs. without contrast, US with contrast was associated with sensitivity of 0.89 (95% CI, 0.83 to 0.93) and US without contrast with a sensitivity of 0.39 (95% CI, 0.32 to 0.47), for a difference in sensitivity of 0.50 (95% CI, 0.41 to 0.58). Based on across-study comparisons, there were no clear differences in sensitivity between different US contrast agents; no study directly compared different contrast agents. There were no differences in sensitivity of US based on lesion depth (3 studies) or body mass index (2 studies).
Other factors	CT	Low	Evidence on effects of technical parameters (type of CT scanner, use of delayed phase imaging, section thickness) was limited by small numbers of studies with wide CIs and methodological limitations, precluding reliable conclusions. Two studies found no clear difference in sensitivity of CT for HCC in patients with vs. without cirrhosis.
Other factors	MRI	Low	There were no clear differences in estimates of sensitivity based on the type of MRI machine (3.0 T vs. 1.5 T), type of contrast, use of delayed phase imaging, timing of delayed phase imaging, and section thickness. Estimates were similar when studies that used diffusion-weighted imaging were excluded.

CI = confidence interval; CT = computed tomography; HCC = hepatocellular carcinoma; MRI = magnetic resonance imaging; US = ultrasound.

Table J. Summary of evidence for Key Question 2.b (evaluation of focal liver lesions): clinical decisionmaking

Imaging Modality or Comparison	Strength of Evidence	Summary
All	Insufficient	No evidence

Table K. Summary of evidence for Key Question 2.c (evaluation of focal liver lesions): clinical and patient-centered outcomes

Imaging Modality or Comparison	Strength of Evidence	Summary
All	Insufficient	No evidence

Table L. Summary of evidence for Key Question 2.d (evaluation of focal liver lesions): harms

Imaging Modality or Comparison	Strength of Evidence	Summary
US and CT	Insufficient	One study of US (with and without contrast) and CT reported harms, but did not stratify results by imaging technique. The overall rate of adverse drug-related events was 10%, with all events classified as mild.

CT = computed tomography; US = ultrasound.

Table M. Summary of evidence for Key Question 3.a (staging): test performance

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Staging accuracy, using TNM criteria	CT	Moderate	The proportion correctly staged using TNM or BCLC criteria ranged from 28% to 58%, the proportion overstaged from 2% to 27%, and the proportion understaged from 25% to 52%, based on 6 studies.
Staging accuracy, using TNM criteria	MRI	Low	The proportion correctly staged ranged from 40% to 75%, the proportion overstaged from 3.1% to 31%, and the proportion understaged from 19% to 31%, based on 3 studies.
Staging accuracy, using TNM criteria	PET	Low	One study found 26% of patients were correctly staged with FDG PET and 91% with 11C-choline PET.
Staging accuracy, using TNM criteria	MRI vs. CT	Low	Two studies reported similar staging accuracy.
Identification of metastatic disease *Unit of analysis: patients with metastatic HCC*	PET	Sensitivity: Low Specificity: Low	Sensitivity of FDG PET was 0.85 (95% CI, 0.71 to 0.93; 6 studies) and specificity 0.93 (95% CI, 0.89 to 0.95; 5 studies), for an LR+ of 11 (95% CI, 7.8 to 17) and LR- of 0.16 (95% CI, 0.08 to 0.33). One study that directly compared sensitivity of FDG PET vs. 11C-acetate PET reported comparable sensitivity (0.79 vs. 0.71), although sensitivity was higher when both tracers were used (0.98).
Identification of metastatic disease *Unit of analysis: patients with metastatic HCC*	PET/CT vs. CT	Low	Three studies found no difference in sensitivity (0.82; 95% CI, 0.61 to 0.93 vs. 0.85; 95% CI, 0.66 to 0.95).
Identification of metastatic disease *Unit of analysis: metastatic HCC lesions*	PET	Sensivity: Low Specificity: Insufficient	Sensitivity of FDG PET was 0.82 (95% CI, 0.72 to 0.90; 5 studies). One study that directly compared sensitivity of FDG vs. 11C-acetate PET reported comparable sensitivity (0.86 vs. 0.77).

BCLC = Barcelona Clinic Liver Cancer; CI = confidence interval; CT = computed tomography; FDG = 18F-fluorodeoxyglucose; HCC = hepatocellular carcinoma; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; MRI = magnetic resonance imaging; PET = positron emission tomography; TNM = tumor, node, metastasis staging.

Table N. Summary of evidence for Key Question 3.a.i (staging): effects of reference standard on test performance

Imaging Modality or Comparison	Strength of Evidence	Summary
CT, MRI, PET	Sensitivity: Insufficient Specificity: Insufficient	Evidence was insufficient to determine effects of different reference standards on accuracy of staging using TNM criteria or accuracy of PET for identifying metastatic HCC because few studies evaluated alternative reference standards.

CT = computed tomography; HCC = hepatocellular carcinoma; MRI = magnetic resonance imaging; PET = positron emission tomography; TNM = tumor, node, metastasis staging.

Table O. Summary of evidence for Key Question 3.a.ii (staging): effects of patient, tumor, and technical factors on test performance

Imaging Modality or Comparison	Strength of Evidence	Summary
CT, MRI, PET	Insufficient	For accuracy of staging using TNM criteria, no study evaluated effects of patient-level characteristics or other factors on accuracy of imaging techniques for staging.
PET	Low	In 1 study that directly compared sensitivity of PET vs. PET/CT for identifying metastatic HCC lesions, there was no clear difference in sensitivity. Four studies of FDG PET found sensitivity increased as lesion size increased, but the number of lesions <1 cm was small (total of 20). Eight studies generally found sensitivity of FDG PET higher for lymph and bone metastasis than for lung metastasis, but samples were small, precluding strong conclusions.

CT = computed tomography; FDG = 18F-fluorodeoxyglucose; HCC = hepatocellular carcinoma; MRI = magnetic resonance imaging; PET = positron emission tomography; TNM = tumor, node, metastasis staging.

Table P. Summary of evidence for Key Question 3.b (staging): clinical decisionmaking

Subquestion	Imaging Modality or Comparison	Strength of Evidence	Summary
Transplant eligibility, using Milan criteria	CT	Moderate	The proportion correctly assessed for transplant eligibility ranged from 40% to 96%. The proportion of patients who met transplant criteria based on CT but exceeded criteria based on the reference standard was 3.5% to 7.8%, based on 3 studies. Two studies found that 2.3% and 16% of patients who underwent transplantation based on Milan criteria had no HCC lesions on examination of explanted livers.
Transplant eligibility, using Milan criteria	CT vs. MRI	Low	One study reported similar accuracy.
Transplant eligibility, using Milan criteria	PET vs. CT	Low	One study found 11C-choline PET more accurate than CT (95% vs. 40%).
Use of resection and ablative therapies	MRI vs. CT	Low	One study reported that the proportion of decisions to perform resection or ablative therapies that were classified as correct were similar for MRI (90% and 90%, respectively) and CT (80% and 77%, respectively).

CT = computed tomography; HCC = hepatocellular carcinoma; MRI = magnetic resonance imaging; PET = positron emission tomography.

Table Q. Summary of evidence for Key Question 3.c (staging): clinical and patient-centered outcomes

Imaging Modality or Comparison	Strength of Evidence	Summary
US with contrast vs. US without contrast plus CT	Low	One cohort study found that contrast-enhanced US identified more small (≤2 cm) HCC lesions than noncontrast US plus CT (36 vs. 31) and was associated with a higher complete necrosis rate following ablation (92%, or 106/115, vs. 83%, or 93/112 lesions; p = 0.036) but was rated high risk of bias. Another study that appeared to be performed in the same series of patients found US with contrast prior to radiofrequency ablation associated with lower local tumor progression rate (7.2% vs. 18%; rate ratio, 0.40; 95% CI, 0.16 to 0.87) and longer tumor-free survival (38 vs. 26 months), but was also rated high risk of bias.

CI = confidence interval; CT = computed tomography; HCC = hepatocellular carcinoma; US = ultrasound.

Table R. Summary of evidence for Key Question 3.d (staging): harms

Imaging Modality or Comparison	Strength of Evidence	Summary
All	Insufficient	No evidence

Findings in Relationship to What Is Already Known

Unlike our review, several previously published reviews on detection of HCC and evaluation of focal liver lesions found no clear differences in test performance between US, CT, and MRI for HCC.[25-28] Several factors may explain these discrepancies: we included more studies than any prior review, separately analyzed studies based on the reason for imaging, stratified studies according to the unit of analysis, and focused on within-study (direct) comparisons of two or more imaging modalities against a common reference standard instead of relying primarily or solely on across-study (indirect) estimates of test performance. Our review's findings are consistent with those of previous reviews regarding lower sensitivity of imaging for detection of small and well-differentiated HCC lesions.

Our findings regarding test performance of PET for detection of metastatic HCC are consistent with those from a recently published systematic review and meta-analysis.[29] Like our review, a recent systematic review found insufficient evidence to determine effects of surveillance with imaging on clinical outcomes.[30] A systematic review on screening for HCC in chronic liver disease funded by the U.S. Department of Veterans Affairs was conducted at the same time as our review.[12]

Applicability

A number of potential issues could impact the applicability of our findings. Over half of the studies were conducted in Asia, where the prevalence, underlying causes, course, evaluation, and management of chronic liver disease may be different than in the United States. To mitigate potential effects of study country on applicability, we excluded invasive imaging techniques not typically used in the United States, such as CT arterial portography and CT hepatic arteriography, as well as imaging techniques considered inadequate in the United States (such as C-arm CT). We also performed stratified analyses focusing on studies performed in the United States and Europe to evaluate effects on estimates of diagnostic accuracy and found no clear effects on estimates.

Imaging techniques are rapidly evolving, which is another factor that could affect applicability. To mitigate effects of outdated techniques on applicability, we excluded imaging technologies considered outdated, such as MRI with magnetic field strength <1.5 T and nonspiral CT, and included only studies published since 1998. We also performed additional analyses on technical factors such as contrast rate, imaging phases evaluated, timing of imaging phases, section thickness, use of hepatobiliary contrast (for MRI), use of diffusion-weighted imaging, and newer technologies (e.g., dual-source or spectral CT). We included studies of US with microbubble contrast even though no agent is currently approved for abdominal imaging in the United States, because efforts to obtain U.S. Food and Drug Administration approval are ongoing and this technique is commonly used in other areas of the world, including Canada and Europe.

As noted above, few studies were performed in true surveillance settings (i.e., in patients at high risk for HCC but not previously diagnosed with this condition). Rather, most studies of test performance that were not performed specifically to evaluate or characterize previously identified lesions were conducted in patients undergoing imaging for other reasons, including series of patients undergoing liver transplantation, surgical resection, or other treatments for HCC. Although such studies are likely to provide some useful findings regarding diagnostic accuracy, results may not be directly applicable to patients undergoing surveillance. In particular, the high prevalence of HCC (many studies enrolled only patients with HCC) could overestimate test performance in true surveillance settings, in which the prevalence of HCC would be much lower.[31]

Implications for Clinical and Policy Decisionmaking

Our review has important potential implications for clinical and policy decisionmaking. Due to the lack of direct evidence regarding clinical benefits and downstream harms associated with different imaging tests for surveillance, diagnosis, and staging of HCC, most decisions regarding use of imaging tests must necessarily be made primarily on the basis of diagnostic test performance. Despite limited

evidence in true surveillance settings, current guidelines from the AASLD recommend US without contrast for surveillance of HCC in at-risk individuals. Evidence from true surveillance settings to evaluate the comparative test performance of different imaging modalities was limited. Based primarily on studies conducted in nonsurveillance settings, our study suggests that US without contrast is less sensitive than MRI or CT for detecting HCC.[4] Although sensitivity for identifying HCC was higher for CT and MRI than for US in studies conducted in nonsurveillance settings, findings may not be directly applicable to clinical and policy decisions related to surveillance, as the spectrum of patients evaluated in these studies could have affected estimates.

In patients found to have an HCC lesion on surveillance, our review supports use of CT and MRI to further characterize lesions >1 cm in size, as in the AASLD guideline, based on high sensitivity and specificity. Evidence is limited but appears consistent with the sequential diagnostic imaging algorithm as outlined in the AASLD guideline, in which typical findings for HCC on sequentially performed CT or MRI are considered sufficient to make a diagnosis.

Our findings also support minimal technical specifications for MRI and CT for HCC imaging, as suggested in recent guidance, such as those regarding minimum contrast rates and use of delayed phase imaging.[11] Evidence suggesting superior test performance of MRI with hepatic-specific versus nonhepatic-specific contrast appears promising, although differences were relatively small. Therefore, clinical and policy decisions around use of nonhepatic-specific contrast may be impacted by additional factors other than test performance, such as cost, harms, or convenience.

US with contrast was associated with test performance similar to that of MRI and CT for evaluation of lesions, although no microbubble contrast agents are currently approved for use in the United States. Although the role of PET is likely to remain focused on identification of metastatic HCC and staging, additional research could help clarify the role of PET with alternative tracers for identification and evaluation of intrahepatic HCC.

Research Gaps

Significant research gaps limit the full understanding of the comparative effectiveness of imaging for surveillance, diagnosis, and staging of HCC. The only randomized trial of effects of surveillance for HCC with imaging on clinical outcomes had important methodological shortcomings and was performed in China, potentially limiting applicability to screening in the United States.[24] Although conducting a randomized trial of surveillance versus no screening in the United States could be difficult because screening is recommended in clinical practice guidelines and routinely performed in high-risk patients, randomized trials that compare screening using different imaging modalities or combinations of modalities would be helpful for understanding optimal approaches.

In lieu of such studies, evidence on effects of alternative imaging strategies on intermediate outcomes such as clinical decisionmaking, subsequent procedures, and resource utilization could also be informative. Such studies could potentially enroll smaller samples than randomized trials to compare screening using different imaging modalities and would probably not require the extended followup needed to assess clinical outcomes.

Although many studies are available on test performance of alternative imaging modalities and strategies, important research gaps remain. Notably, few studies evaluated imaging in true surveillance settings, and evidence on accuracy of imaging for identifying HCC lesions from nonsurveillance settings may not be directly applicable to surveillance due to spectrum effects. More studies are also needed to clarify the role of promising alternative techniques, such as US with contrast, MRI with hepatic-specific contrast, and PET with alternative tracers, on estimates of accuracy. Research should focus on improving methods for identifying small or well-differentiated HCC lesions, for which imaging remains suboptimal.

Conclusions

Based on estimates of test performance, several imaging modalities appear to be reasonable options for detection of HCC, evaluation of focal liver lesions for HCC, or staging of HCC. Although there are some potential differences in test performance between different imaging modalities and techniques, more research is needed to understand the effects of such differences on clinical decisionmaking and clinical outcomes.

References

1. Parkin DM. Global cancer statistics in the year 2000. Lancet Oncol. 2001 Sep;2(9):533-43. PMID: 11905707.
2. National Cancer Institute. Liver Cancer. www.cancer.gov/cancertopics/types/liver. Accessed September 21, 2011.

3. Howlader N, Noone AM, Krapcho M, et al., eds. SEER Cancer Statistics Review, 1975-2010. National Cancer Institute: Bethesda, MD; 2013. http://seer.cancer.gov/csr/1975_2010/.
4. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. Hepatology. 2011 Mar;53(3):1020-2. PMID: 21374666.
5. Cabibbo G, Enea M, Attanasio M, et al. A meta-analysis of survival rates of untreated patients in randomized clinical trials of hepatocellular carcinoma. Hepatology. 2010 Apr;51(4):1274-83. PMID: 20112254.
6. Forner A, Reig ME, de Lope CR, et al. Current strategy for staging and treatment: the BCLC update and future prospects. Semin Liver Dis. 2010 Feb;30(1):61-74. PMID: 20175034.
7. Mazzaferro V, Bhoori S, Sposito C, et al. Milan criteria in liver transplantation for hepatocellular carcinoma: an evidence-based analysis of 15 years of experience. Liver Transpl. 2011;17(S2):S44-S57. PMID: 21695773.
8. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2014. Chapters available at www.effectivehealthcare.ahrq.gov.
9. Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EHC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
10. Submit Scientific Information Packets. Rockville, MD: Agency for Healthcare Research and Quality; 2013. www.effectivehealthcare.ahrq.gov/index.cfm/submit-scientific-information-packets/. Accessed November 12, 2013.
11. Wald C, Russo MW, Heimbach JK, et al. New OPTN/UNOS policy for liver transplant allocation: standardization of liver imaging, diagnosis, classification, and reporting of hepatocellular carcinoma. Radiology. 2013 Feb;266(2):376-82. PMID: 23362092.
12. Kansagara D, Papak J, Pasha A, et al. Screening for Hepatocellular Cancer in Chronic Liver Disease: A Systematic Review. Washington: Department of Veterans Affairs; January 2014. www.ncbi.nlm.nih/gov/books/NBK2221841.
13. Harris RP, Helfand M, Woolf SH, et al. Current methods of the U.S. Preventive Services Task Force: a review of the process. Am J Prev Med. 2001 Apr;20(3 Suppl):21-35. PMID: 11306229.
14. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36. PMID: 22007046.
15. Fu R, Gartlehner G, Grant M, et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Chapters available at www.effectivehealthcare.ahrq.gov/.
16. Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. Ann Intern Med. 2013 Apr 2;158(7):544-54. PMID: 23546566.
17. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011 Nov;64(11):1198-207. PMID: 21463926.
18. Chalasani N, Horlander JC Sr, Said A, et al. Screening for hepatocellular carcinoma in patients with advanced cirrhosis. Am J Gastroenterol. 1999 Oct;94(10):2988-93. PMID: 10520857.
19. Van Thiel DH, Yong S, Li SD, et al. The development of de novo hepatocellular carcinoma in patients on a liver transplant list: frequency, size, and assessment of current screening methods. Liver Transpl. 2004 May;10(5):631-7. PMID: 15108254.
20. Goto E, Masuzaki R, Tateishi R, et al. Value of post-vascular phase (Kupffer imaging) by contrast-enhanced ultrasonography using Sonazoid in the detection of hepatocellular carcinoma. J Gastroenterol. 2012;47(4):477-85. PMID: 22200940.
21. Kunishi Y, Numata K, Morimoto M, et al. Efficacy of fusion imaging combining sonography and hepatobiliary phase MRI with Gd-EOB-DTPA to detect small hepatocellular carcinoma. AJR Am J Roentgenol. 2012 Jan;198(1):106-14. PMID: 22194485.
22. Lencioni R, Piscaglia F, Bolondi L. Contrast-enhanced ultrasound in the diagnosis of hepatocellular carcinoma. J Hepatol. 2008 May;48(5):848-57. PMID: 18328590.
23. Naaktgeboren CA, de Groot JA, van Smeden M, et al. Evaluating diagnostic accuracy in the face of multiple reference standards. Ann Intern Med. 2013 Aug 6;159(3):195-202. PMID: 23922065.
24. Zhang BH, Yang BH, Tang ZY. Randomized controlled trial of screening for hepatocellular carcinoma. J Cancer Res Clin Oncol. 2004 Jul;130(7):417-22. PMID: 15042359.
25. Colli A, Fraquelli M, Casazza G, et al. Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review. Am J Gastroenterol. 2006 Mar;101(3):513-23. PMID: 16542288.
26. Westwood M, Joore M, Grutters J, et al. Contrast-enhanced ultrasound using SonoVue(R) (sulphur hexafluoride microbubbles) compared with contrast-enhanced computed tomography and contrast-enhanced magnetic resonance imaging for the characterisation of focal liver lesions and detection of liver metastases: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2013 Apr;17(16):1-243. PMID: 23611316.
27. Guang Y, Xie L, Ding H, et al. Diagnosis value of focal liver lesions with SonoVue-enhanced ultrasound compared with contrast-enhanced computed tomography and contrast-enhanced MRI: a meta-analysis. J Cancer Res Clin Oncol. 2011 Nov;137(11):1595-605. PMID: 21850382.
28. Xie L, Guang Y, Ding H, et al. Diagnostic value of contrast-enhanced ultrasound, computed tomography and magnetic resonance imaging for focal liver lesions: a meta-analysis. Ultrasound Med Biol. 2011 Jun;37(6):854-61. PMID: 21531500.
29. Lin C-Y, Chen J-H, Liang J-A, et al. 18F-FDG PET or PET/CT for detecting extrahepatic metastases or recurrent hepatocellular carcinoma: a systematic review and meta-analysis. Eur J Radiol. 2012 Sep;81(9):2417-22. PMID: 21899970.

30. Aghoram R, Cai P, Dickinson JA. Alpha-foetoprotein and/or liver ultrasonography for screening of hepatocellular carcinoma in patients with chronic hepatitis B. Cochrane Database Syst Rev. 2012;(9):CD002799. PMID: 22972059.

31. Lachs MS, Nachamkin I, Edelstein PH, et al. Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. Ann Intern Med. 1992 Jul 15;117(2):135-40. PMID: 1605428.

Full Report

This executive summary is part of the following document: Chou R, Cuevas C, Fu R, Devine B, Wasson N, Ginsburg A, Zakher B, Pappas M, Graham E, Sullivan S. Imaging Techniques for the Diagnosis and Staging of Hepatocellular Carcinoma. Comparative Effectiveness Review No. 143. (Prepared by the Pacific Northwest Evidence-based Practice Center under Contract No. 290-2012-00014-I.) AHRQ Publication No. 14(15)-EHC048-EF. Rockville, MD: Agency for Healthcare Research and Quality; October 2014. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

AHRQ Pub. No. 14(15)-EHC048-1-EF
October 2014